Ref. Ares(2025)8986171 - 21/10/2025

Musical Metaverse made in Europe: An innovation lab for musicians and audiences of the future

D3.1: Publication on 5G potentialities and limitations for the Musical Metaverse

European Innovation Council
Funded by the European Union

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101184379. This work was supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 25.00026.

Schweizerische Eidgenossenschaft
Confédération suisse
State Secretariat for Education, Research and Innovation SERI

Document Information

Document factsheet

Work Package	WP 3 ML over Wireless Networks
Task	T3.1 Communication Strategy
Due date	31/10/2025
Deliverable lead	KTH
Authors	Krishnendu S. Tharakan, Carlo Fischione, Omran Ayoub, Ali Alhousseini
Reviewers	Francesco Malandrino, Luca Turchet
Nature of deliverable	R
Dissemination level	PU

Document history

Version	Date	Description
v1.0	31/07/2025	Draft creation
v1.1	10/10/2025	First revision
v1.2	17/10/2025	Second revision
v1.3	21/10/2025	Document finalized

Disclaimer

This project has received funding from The European Union as part of the Horizon Europe Framework Program (HORIZON) under grant agreement No. 101184379. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Innovation Council. Neither the European Union nor the European Innovation Council can be held responsible for them.
This work was also supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 25.00026.

Table of contents

LIST OF ACRONYMS **1**

EXECUTIVE SUMMARY **2**

1. INTRODUCTION **3**

1.1 Context and Scope of the Musical Metaverse 3
1.2 Potentialities of 5G for the Musical Metaverse 3
1.3 Limitations of Current 5G Deployments for MM 4

2. TECHNICAL ANALYSIS OF 5G FOR MM **5**

2.1 Network Slicing and MM Slice Design 5
2.2 Low-Latency and Reliable Transmission of Multimodal Musical Data 6
2.3 MEC and Edge AI for MM Applications 6
2.4 Digital Twin Integration in the MM 7
2.5 Reinforcement Learning for Adaptive Orchestration 7

3. POTENTIALITIES IN DETAIL **8**

3.1 Ultra-Reliable Low-Latency Communication (URLLC) for MM 8
3.2 Enhanced Mobile Broadband (eMBB) 8
3.3 Massive Machine-Type Communications (mMTC) for IoMusT Integration 8
3.4 Enhanced Mobile Broadband (eMBB) for Immersive Experiences 9

4. LIMITATIONS AND TECHNICAL BOTTLENECKS **10**

4.1 Public Non-Standalone (NSA) Constraints 10
4.2 Incomplete Slicing Implementation 10
4.3 Uplink Stability and Multi-User Synchronization Issues 10
4.4 MEC Deployment Gaps 10
4.5 Interoperability Challenges in Vendor Equipment 11
4.6 Sensor Density and Communication Overhead Trade-offs 11
4.7 Standardization and Governance Gaps 11

5. SIMULATOR **12**

5.1 Limitations of classical VNE for Musical Metaverse Requirements 12
5.1.1 Cloud network graph modeling *13*
5.1.2 Service Graph Modeling *13*
5.2 *MuMeNet* simulator 14
5.2.1 MuMeNet vs existing Network Simulators *15*
5.2.2 MuMeNet: Architecture and Design *15*

6. FUTURE DEVELOPMENTS AND RESEARCH DIRECTIONS **19**

7. CONCLUSION **20**

List of acronyms

5GC	5G Core
5G NSA	5G Non-Standalone
5G SA	5G Standalone
CNFlow	Cloud Network Flow
CNG	Cloud Network Graph
DSPaaS	Digital Signal Processing as a Service
DT	Digital Twin
EEG	Electroencephalogram
eMBB	enhanced Mobile Broadband
EPC	Evolved Packet Core
HRIR	Head-Related Impulse Responses
IoMusT	Internet of Musical Things
MARL	Multi-Agent Reinforcement Learning
MEC	Multi-access Edge Computing
MM	Musical Metaverse
mmWave	Millimeter-wave
MR	Mixed Reality
mMTC	Machine-Type Communications
NMP	Networked Music Performance
PLC	Packet Loss Concealment
QoE	Quality of Experience
QoS	Quality of Service
RL	Reinforcement Learning
SG	Service Graph
URLLC	Ultra-Reliable Low-Latency Communication
VNE	Virtual Network Embedding
VNR	Virtual-Network Request
XR	Extended Reality

Executive Summary

This deliverable presents the findings from our investigation into the capabilities and constraints of 5G networks in supporting the Musical Metaverse (MM), a domain that integrates networked music performance (NMP), immersive extended reality environments, and intelligent orchestration to enable real-time, interactive musical experiences across global distances. The deliverable also introduces a novel network simulator specific for MM applications.

The content builds on our articles *"Towards a Networking Architecture for the Musical Metaverse: The MUSMET Vision"*, *"Architecting the Musical Metaverse: Lessons from 5G and Emerging Technologies"* and *"MuMeNet: A Network Simulator for Musical Metaverse Communications"*, all accepted for publication in the 6th IEEE International Symposium on Internet of Sounds (publicly released as green open access on the project website). In addition to summarizing the papers' findings, this deliverable expands them with a forward-looking research roadmap to guide further development within the MUSMET project.

Our analysis confirms that 5G's key features, including network slicing and Multi-access Edge Computing (MEC), can meet many of the stringent demands of MM applications, that match those of Ultra-Reliable Low-Latency Communication, one of the main 5G use cases. Experimental evidence shows that private Standalone 5G deployments with MEC integration can sustain <30 ms round-trip latency with high reliability, enabling synchronous audio, video, haptic, and sensor data exchange required for NMP.

However, current public Non-Standalone deployments and partial slicing implementations still suffer from latency spikes, jitters, uplink instability, and limited edge computing availability. These bottlenecks are particularly problematic in multi-user MM sessions and when streaming multimodal data at scale. This deliverable outlines a future development plan centered on AI-driven orchestration, deeper digital twin integration, advanced connectivity paradigms beyond 5G (e.g., 6G, THz, Reconfigurable Intelligent Surfaces), and strong standardization for interoperability.

1. Introduction

This deliverable presents a set of findings resulting from the collaboration between researchers from KTH Royal Institute of Technology, University of Applied Sciences and Arts of Southern Switzerland, Politecnico di Torino, Consiglio Nazionale delle Ricerche, and University of Trento.

The first part of the deliverable describes the investigation into the potentialities and limitations of 5G networks in supporting the Musical Metaverse (MM), a domain that integrates networked music performance (NMP), immersive extended reality (XR) environments, and intelligent orchestration to enable real-time, interactive musical experiences across global distances. The second part of the deliverable introduces a novel network simulator specific for MM applications.

The deliverable builds upon the following three publications, whose preprint is made publicly available as green open access on the project website (https://musmet.eu/publications/):

- L. Turchet, A. Boem, O. Ayoub, F. Malandrino, J. Llorca, C. Fischione and C. Rottondi. *Towards a Networking Architecture for the Musical Metaverse: The MUSMET Vision*. In Proceedings of the 6th IEEE International Symposium on the Internet of Sounds, 2025.
- C. Rinaldi, K. S. Tharakan, L. Turchet, C. Rottondi, C. Centofanti and C. Fischione. *Architecting the Musical Metaverse: Lessons from 5G and Emerging Technologies*. In Proceedings of the 6th IEEE International Symposium on the Internet of Sounds, 2025.
- Al Housseini, J. Llorca, L. Turchet, T. Leidi, C. Rottondi and O. Ayoub. *MuMeNet: A Network Simulator for Musical Metaverse Communications*. In Proceedings of the 6th IEEE International Symposium on the Internet of Sounds, 2025.

1.1 Context and Scope of the Musical Metaverse

The MM extends the concept of Metaverse to music, creating persistent, interactive, and immersive spaces where musicians, audiences, and AI agents interact in real time. It is built upon the Internet of Musical Things (IoMusT), which connects smart instruments, wearables, and performance systems into a distributed, networked ecosystem.

MM scenarios range from live extended reality (XR) concerts and distributed jam sessions to collaborative composition environments, requiring precise synchronization of multimodal data: audio, video, haptic cues, gestures, and even EEG signals. These requirements push wireless networks to their operational limits, making 5G a key enabler.

5G has emerged as a foundational technology for this vision, offering Ultra-Reliable Low-Latency Communication (URLLC), flexible network slicing, and multi-access edge computing (MEC) integration. However, musical applications represent one of the most demanding use cases for wireless networks, pushing the limits of both radio access and core network performance. This deliverable investigates how far current 5G deployments can meet these demands, identify technical bottlenecks, and suggest pathways for overcoming them.

1.2 Potentialities of 5G for the Musical Metaverse

5G technologies have been designed and deployed to serve three main families of use cases, namely, URLLC, enhanced Mobile Broadband (eMBB), and Machine-Type Communications (mMTC). The requirements of these use cases match those of the MM, which makes 5G a key enabler for the MM itself. Specifically, as required for URLLC, 5G allows transmission times below one millisecond at the physical layer, which is essential for achieving the sub-30 millisecond end-to-end latency required in professional NMP. Through network slicing, dedicated “MM slices” can be created and optimized for specific data types, such as audio, XR video, and sensor information effectively minimizing interference

from other traffic. MEC further enhances system responsiveness by enabling local processing of latency-critical tasks, including audio rendering, gesture recognition, and XR scene updates.

Additionally, as required for eMBB, 5G provides high throughput to support simultaneous high-definition or 4K video and spatial audio streams for large numbers of participants. Finally, as in the mMTC use case, 5G allows the integration of dense networks of sensors, enabling detailed motion capture, haptic feedback, and environmental awareness, key to creating immersive and expressive musical experiences in the Metaverse.

1.3 Limitations of Current 5G Deployments for MM

Despite the significant potential of 5G, several limitations remain in current deployments that hinder the realization of large-scale MM applications. Public Non-Standalone (NSA) networks still rely on 4G core components, which introduce additional latency and jitter, making them unsuitable for ultra-low-latency performance scenarios. Moreover, many commercial 5G implementations lack complete and dynamic network slicing capabilities, preventing the creation of programmable slices with real-time prioritization for critical audio or XR data streams. Uplink stability also remains a challenge, as public networks often experience fluctuations that disrupt multi-source audio transmissions typical in collaborative performances.

The deployment of MEC is another constraint, its availability is geographically limited, leading to inconsistent performance and increased dependence on centralized cloud resources. Interoperability issues between different vendors' equipment and software further restrict end-to-end optimization and integration of advanced MM services. Finally, while higher sensor density improves precision in motion capture and environmental awareness, it also increases communication overhead, demanding careful trade-offs between sensing accuracy, latency, and bandwidth efficiency.

2. Technical Analysis of 5G for the MM

2.1 Network Slicing and MM Slice Design

Network slicing enables the logical partitioning of physical 5G infrastructure into virtual network instances tailored to specific service requirements. In the context of the MM, slicing facilitates differentiated handling of multimodal data flows, audio, video, haptic, and sensor information, each with distinct latency and bandwidth needs. Dedicated "MM slices" can be designed to prioritize ultra-low-latency audio and haptic traffic while allocating separate channels for less time-critical visual rendering or metadata transmission.

Empirical results from private 5G deployments show that properly configured slices can reduce packet loss by up to 20% and maintain sub-25 ms end-to-end latency even under high traffic load. However, orchestration overhead from slice isolation can marginally increase baseline latency (typically by 0.2–0.3 ms). The next step in slicing research for MM involves AI-based dynamic orchestration, where reinforcement learning agents monitor network conditions and reconfigure slices in real time to optimize resource allocation without manual intervention.

The development of MM-specific slice templates, integrating spatial audio, motion tracking, and real-time XR rendering, represents a crucial step toward establishing reusable architectures for musical applications within next-generation mobile networks.

Fig. 1 illustrates a functional stack of enabling technologies for the MM, highlighting the layered orchestration of communication, computation, intelligence, and immersive interaction.

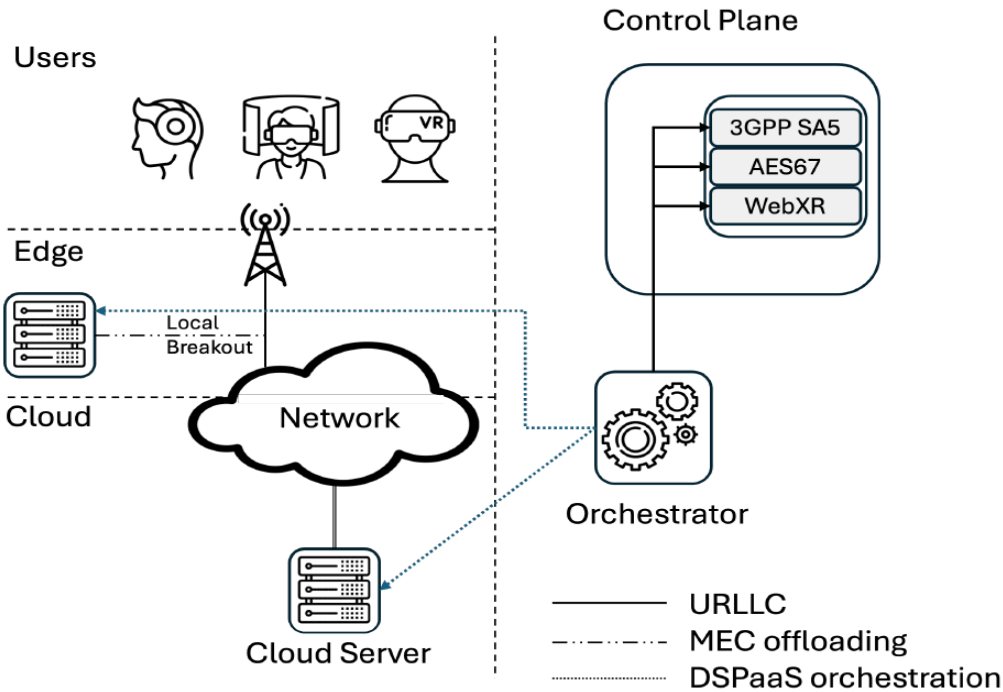

Figure 1. *Functional stack of enabling technologies for the Musical Metaverse, highlighting the layered orchestration of communication, computation, intelligence, and immersive interaction.*

2.2 Low-Latency and Reliable Transmission of Multimodal Musical Data

Multimodal synchronization is central to the MM experience. Real-time interaction requires seamless integration of audio, visual, gestural, haptic, and physiological data (e.g., EEG). Audio transmission demands sub-30 ms round-trip latency, while XR video can tolerate higher delays (up to 100 ms). The challenge lies in aligning these modalities without perceptual degradation. Studies on 5G-enabled NMP systems confirm that even a 20 ms inter-stream offset between audio and visual data can disrupt the sense of synchrony and immersion.

Deep learning-based Packet Loss Concealment (PLC) algorithms have emerged as an effective solution to address transient network instabilities. Hybrid autoregressive deep learning PLC models, such as PARCnet, have demonstrated superior performance, achieving real-time concealment on CPUs with perceptual scores close to studio-grade references. These methods can be extended to gesture and video data streams to maintain cross-modal coherence during packet loss or delay spikes.

Moreover, joint optimization frameworks using multi-objective reinforcement learning (e.g., Multi-Objective Soft Actor-Critic) have shown promise for coordinating bandwidth, encoding, and task offloading, ensuring a balanced Quality of Service (QoS) and Quality of Experience (QoE) across multimodal streams.

Fig. 2 provides an overview of the general networking architecture envisioned for the project.

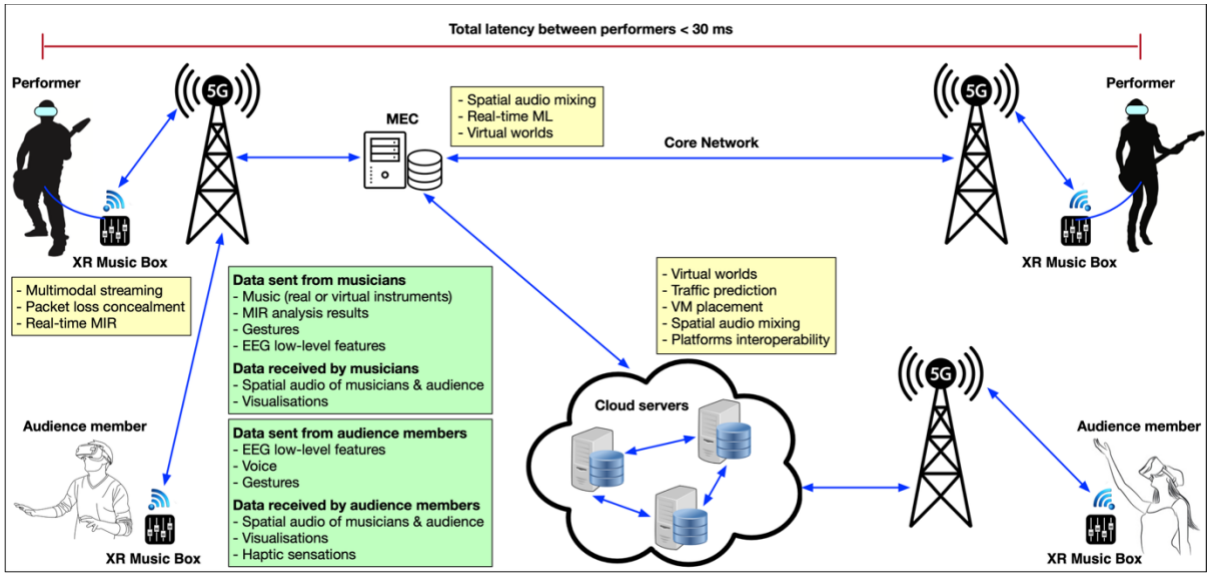

Figure 2. *The MUSMET general architecture.*

2.3 MEC and Edge AI for MM Applications

MEC represents a cornerstone for real-time MM experiences by reducing round-trip time and decentralizing computational loads. In MM scenarios, MEC nodes perform latency critical operations such as real-time spatial audio rendering (using HRIR convolution or Ambisonics), dynamic room acoustic modeling, and gesture recognition. Edge based deployment of these workloads can reduce end-to-end latency by 30–40% compared to centralized cloud architectures.

When combined with AI models, MEC transforms from a passive offloading point into an intelligent orchestration layer. For example, edge-deployed neural networks can perform adaptive bitrate control, predictive buffering, and resource allocation based on real-time traffic conditions. However, MEC

availability is uneven across geographic regions, and heterogeneity in hardware (e.g., GPUs vs. CPUs) affects consistency in AI inference performance.

The integration of "Digital Signal Processing as a Service" (DSPaaS) over MEC further extends this paradigm, allowing modular, on-demand deployment of audio effects chains that can be dynamically instantiated and orchestrated during performances. Future MM infrastructures are expected to combine MEC and AI in hybrid cloud-edge frameworks, ensuring both low-latency interactivity and global scalability.

2.4 Digital Twin Integration in the MM

Digital Twins (DTs) serve as real-time, adaptive digital counterparts of physical entities such as performers, instruments, and venues. In MM environments, DTs are not merely replicas but predictive models capable of simulating musical and environmental dynamics. For example, a DT of a concert hall can anticipate changes in acoustic response as performers move, while a DT of a musician can forecast expressive gestures, enabling preemptive resource allocation and synchronization.

Modern DT implementations integrate deep learning architectures to process multimodal inputs (motion capture, haptic sensors, and EEG signals), allowing nuanced interpretation of expressive behavior. This integration enables novel applications, such as AI assisted rehearsal planning, cross-reality improvisation, and real-time adaptation of virtual acoustics.

The challenge remains in ensuring ultra-fast synchronization between physical and digital domains, as even small delays in DT updates can disrupt co-creative interactions. Future work envisions multi-agent reinforcement learning enabled DT ecosystems where agents collaboratively optimize network, compute, and perceptual parameters for each performance in real time.

2.5 Reinforcement Learning for Adaptive Orchestration

Reinforcement Learning (RL) and its multi-agent extensions (MARL) provide a promising framework for dynamic orchestration in the MM. RL agents can monitor network metrics, latency, jitter, throughput, as well as user-centric signals like synchronization error and perceptual QoE, adapting system parameters to maximize performance.

In distributed ensemble performances, MARL can coordinate multiple agents (representing performers, network controllers, or edge nodes) to achieve joint optimization of bandwidth allocation, slice prioritization, and processing load distribution. This ensures not only technical stability, but also musical expressivity, as agents can learn to anticipate structural changes in the performance (e.g., crescendos, tempo shifts) and reallocate resources accordingly.

An illustrative scenario involves an RL based edge orchestrator that detects congestion in one slice and proactively migrates audio processing tasks to a nearby MEC node, preventing desynchronization without human intervention. This paradigm moves MM infrastructures from static configurations toward self-learning, adaptive systems that can sustain complex multimodal performances autonomously.

3. Potentialities in Detail

3.1 Ultra-Reliable Low-Latency Communication (URLLC) for MM

URLLC is one of the most critical use cases for 5G, which in turn makes 5G a key enabler for the MM. In NMP, latency must be kept below 25–30 ms round trip to avoid disrupting the rhythmic and harmonic alignment between performers. As required by URLLC, 5G offers physical-layer transmission times as low as 1 ms, combined with high reliability (>99.999%), which ensures that packets containing time-critical audio or control data are delivered promptly and without loss.

In practical MM deployments, 5G allows musicians located in different cities, or even continents to rehearse and perform together in a shared XR stage. This capability extends beyond audio, enabling synchronized haptic feedback for instrument simulation and gesture-based conducting, where delays or jitter would otherwise break immersion. Our experiments confirmed that private standalone (SA) 5G deployments achieve consistent low-latency performance when the network is configured with dedicated slices, while public NSA networks cannot reliably meet the threshold due to additional core network delays and the lack of slicing.

3.2 Enhanced Mobile Broadband (eMBB)

eMBB use cases require 5G to provide multi-gigabit throughput and large spectrum allocations, supporting the bandwidth intensive demands of immersive XR-based music events. A single XR session may combine multiple high-resolution video feeds, 3D holographic projections, and 360° spatial audio streams.

With the support of 5G networks meeting eMBB requirements, it becomes feasible to host virtual music festivals or distributed orchestras where hundreds of participants interact with minimal perceptual delay. Combined with MEC-based rendering, this ensures high-quality audiovisual content with adaptive bitrate control to preserve continuity even under variable network load. Moreover, integrating millimeter-wave (mmWave) technologies within 5G provides sufficient spectrum for parallel HD/4K streams, enabling detailed visualization of performer avatars and realistic virtual venues.

3.3 Massive Machine-Type Communications (mMTC) for IoMusT Integration

The MM relies heavily on the IoMusT, a network of interconnected musical devices such as smart instruments, wearable sensors, MIDI controllers, motion-capture suits, and EEG headsets. These devices continuously generate streams of performance and environmental data. 5G's ability to support up to 1 million devices per km², as foreseen by mMTC use cases, makes it feasible to deploy dense sensor networks in both physical venues and virtual performance hubs. For example, a large-scale virtual concert could feature hundreds of musicians, each with multiple IoMusT devices, transmitting gesture, posture, and acoustic environment data in real time.

The challenge, however, lies in balancing sensor density with network bandwidth. Our findings show that exceeding optimal density thresholds increases synchronization overhead and packet collision rates, necessitating intelligent load-balancing strategies.

3.4 Enhanced Mobile Broadband (eMBB) for Immersive Experiences

The MM is not limited to audio transmission; it also incorporates high-resolution XR visuals, audience interactions, and stage effects. Under eMBB requirements, 5G delivers multi-gigabit speeds, enabling simultaneous 4K/8K video streaming, volumetric video capture, and 3D spatial audio for large audiences. In one use case tested during our experiment, 5G enabled streaming of multi-view camera feeds from performers into a virtual concert hall, allowing audience members to choose their preferred viewing angles in real time. When combined with MEC rendering, this significantly enhanced immersion without introducing noticeable delay. However, the benefits are contingent on consistent uplink performance, which remains a limitation in many public 5G deployments.

4. Limitations and Technical Bottlenecks

4.1 Public Non-Standalone (NSA) Constraints

Most commercial 5G deployments are still NSA, relying on a 4G Evolved Packet Core (EPC) for control signaling. This hybrid configuration adds avoidable latency often 10–20 ms to the user-plane path. Measurements from testbeds reveal that, even under moderate load, public NSA 5G cannot sustain the sub-30 ms end-to-end requirement critical for professional networked-music performance. Jitter fluctuations of 5–15 ms and sporadic latency spikes > 50 ms were recorded when multiple concurrent users shared the same cell, severely impairing synchronization between remote performers. While sufficient for streaming or gaming, NSA architectures cannot deliver deterministic latency or the bidirectional uplink symmetry needed in musical interaction scenarios. The transition to Standalone networks with a native 5G Core (5GC) is therefore mandatory to unlock full control of QoS and URLLC features.

4.2 Incomplete Slicing Implementation

Network slicing, conceptually one of 5G's strongest enablers remains partially implemented in most public infrastructures. Many operators support static, manually configured slices for fixed service categories rather than dynamic, on-demand orchestration based on application context.

For MM use cases, where the traffic pattern changes rapidly between low-bit-rate control data and bursty multimedia streams, such rigidity leads to sub-optimal resource utilization. The absence of cross-domain slice management across radio, transport, and core further limits the ability to enforce consistent latency guarantees. Experimental deployments within research testbeds demonstrate that AI-based orchestration can reduce inter-slice interference and latency variance by 15–20 %, but these mechanisms are not yet standardized. Consequently, most MM sessions must operate on best-effort slices, sacrificing predictability and stability both fatal to ensemble synchronization.

4.3 Uplink Stability and Multi-User Synchronization Issues

The MM reverses the traditional downlink-dominant traffic pattern of mobile networks. In MM sessions, each performer transmits audio, video gestures, and sensor streams simultaneously, producing uplink heavy traffic. Public 5G networks, optimized for video consumption rather than production, suffer from variable uplink throughput and scheduler granularity limitations. Measurements show that uplink bit-rate oscillations of ±25 % cause temporal desynchronization between performers, while buffer underflows trigger audible dropouts. Coordinating multiple distributed performers amplifies the issue, as latency misalignment compounds across peers.

Advanced 5G features such as grant-free uplink access and time-sensitive scheduling could mitigate these effects, but they are not widely available in commercial releases. The lack of symmetrical QoS for uplink traffic remains one of the principal barriers to reliable MM deployment over public networks.

4.4 MEC Deployment Gaps

MEC is essential for reducing round-trip latency and supporting localized computation (audio mixing, spatial rendering, gesture recognition). However, current MEC infrastructure suffers from limited geographical coverage and inconsistent resource provisioning. Operators typically deploy edge

nodes in metropolitan areas, leaving regional and rural zones dependent on distant cloud data centers with > 60 ms latency.

Additionally, MEC hardware heterogeneity (CPU-only vs GPU-accelerated nodes) create variability in AI-inference performance. These inconsistencies complicate the orchestration of time-critical workloads in MM scenarios. For instance, performance spanning multiple cities may experience uneven latency or quality when different performers connect through MEC nodes of unequal capability. The absence of standardized APIs for edge discovery, load balancing, and real-time resource negotiation further exacerbates this challenge.

4.5 Interoperability Challenges in Vendor Equipment

End-to-end optimization in MM requires coordination across devices, access networks, edge nodes, and XR platforms. Current 5G ecosystems, however, are fragmented across vendors and software stacks. Differences in RAN implementations, slicing controllers, and MEC APIs hinder seamless orchestration.

Cross-vendor incompatibility limits the deployment of AI driven resource management frameworks that rely on unified telemetry and control interfaces. As a result, performance tuning becomes site-specific and lacks scalability. Efforts within 3GPP and ETSI MEC working groups are progressing toward open interfaces, but adoption remains uneven. Until interoperability is achieved, MM systems must rely on custom middleware, adding complexity and reducing portability.

4.6 Sensor Density and Communication Overhead Trade-offs

MM experiences depend on dense IoMusT deployments, smart instruments, motion sensors, wearables, and environmental probes generating high-frequency data streams. While increasing sensor density enhances spatial and expressive accuracy, it linearly raises communication overhead. Simulations indicate that doubling sensor density in a virtual venue can increase uplink utilization by 40–50 % and raise synchronization jitter by 6–8 ms without adaptive scheduling. This trade-off constrains scalability for large ensembles or audience-participation events. Adaptive sampling, event-driven data compression, and predictive sensing (using AI models at the edge to infer missing samples) can partially mitigate overhead, but such techniques remain experimental. Balancing sensing fidelity, energy efficiency, and communication stability is therefore an open research problem for large-scale MM deployments.

4.7 Standardization and Governance Gaps

The absence of MM-specific standards hinders interoperability and quality assurance. Existing 5G standards define generic QoS classes but lack mappings to music-specific parameters such as temporal alignment, spectral fidelity, and perceptual QoE. Without standardized interfaces for AI model deployment, DT interoperability, and federated orchestration, each vendor implements proprietary solutions, fragmenting the ecosystem. Cross-disciplinary coordination between telecommunications bodies (3GPP, ETSI) and creative-industry standards (MIDI 2.0, AES SC-02-12) will be vital to achieve sustainable integration.

5. Simulator

5.1 Limitations of classical VNE for Musical Metaverse Requirements

The MM slice place stringent requirements on networking and computing that expose fundamental limitations of traditional models used to describe the slice provisioning problem, such as the classical Virtual Network Embedding (VNE) model. In practice, ensuring end-to-end latencies on the order of only a few milliseconds with near-zero jitter and packet loss is necessary to preserve the illusion of co-presence and maintain musical timing. Moreover, MM scenarios involve multi-modal, interactive media flows beyond just audio. Musicians and audiences exchange not only high-fidelity audio streams, but also real-time control signals, sensor data (e.g., gestures, retrieved emotions, or motion capture), and even shared virtual environment data. Delivering and synchronizing these multimodal streams (audio, video, haptic feedback, etc.) introduces additional challenges in maintaining QoS and QoE. Fig. 3 provides an illustrative example of a musical chunk data stream sent by a musician to two different consumers (listeners).

Classical VNE, however, was not designed with these demands in mind. In the traditional VNE model, a service is abstracted as a virtual network: a set of virtual nodes (virtual machines or functions) connected by virtual links with fixed capacities or bandwidth demands. The goal is to map (embed) this virtual network onto a physical substrate network (infrastructure of physical nodes and links) while respecting resource capacities. VNE assumes a one-to-one mapping: each virtual node is placed on a single substrate node, and each virtual link is assigned a specific path through substrate links. Crucially, classical VNE treats each virtual link as an independent point-to-point traffic demand, without awareness of the content or relationships between flows.

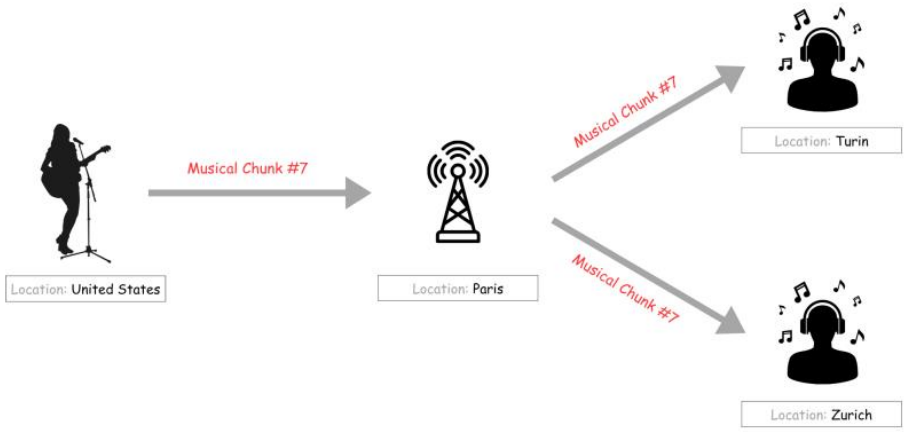

Figure 3. *Illustrative example of a musical chunk data stream sent by a musician to two different consumers (listeners).*

This assumption becomes a critical drawback in the musical metaverse. For example, if one musician's audio stream needs to be delivered to multiple other participants, the VNE model would represent this as multiple separate virtual links carrying identical audio data. The VNE embedding algorithms would attempt to route each link separately, potentially duplicating data over the network and consuming bandwidth on each path redundantly.

To address the shortcomings of VNE, we leverage the Cloud Network Flow (CNFlow) graph-based modeling already introduced for NextG services, and we propose a graph-based modeling that reconceptualizes how services are mapped onto the network graph. In essence, it treats the joint placement of functions, routing of data, and allocation of resources as a single unified information flow

problem on an augmented network graph. Rather than embedding a "virtual network" of nodes and links, this modeling considers the service as a directed graph of information processing (often modeled as a Directed Acyclic Graph, DAG). And the physical infrastructure is a rich cloud–network graph.

5.1.1 Cloud network graph modeling

Unlike VNE, where the substrate network is modeled as an undirected graph that consists of a set of nodes and links. The cloud network graph (CNG) is a directed augmented cloud network graph that represents the network infrastructure as a set of commodities (information flows), allowing for a seamless design for MM scenarios, and enabling studying complex connections between components for data-intensive applications.

The CNG is a directed augmented graph denoted as G = (V, E). Where V represents cloud-network nodes (e.g., core cloud nodes, edge cloud nodes, compute-enabled base stations, or end devices with embedded computing resources), and edges, e ∈ E, represent network links between computing locations (see Fig. 4). Each node u ∈ V, is further augmented as illustrated below, where:

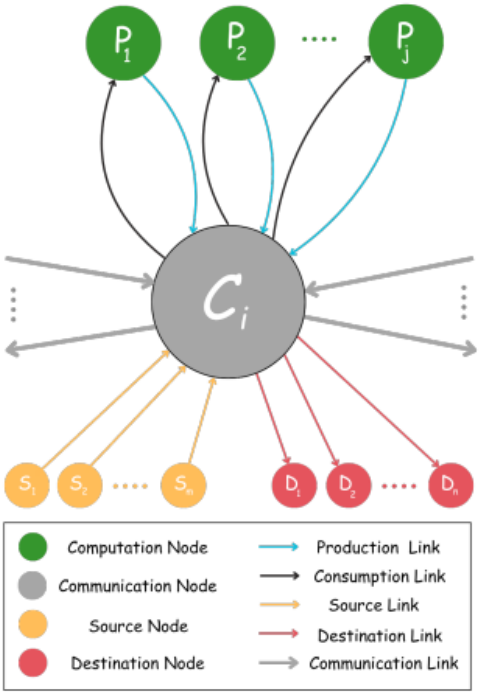

Figure 4. *Augmented node C. Grey edges are inter-site communication links. Colored edges: source, production, consumption, destination.*

- S, D, and P (sets of nodes) and their associated links are used to model *production, consumption* and *processing* of data streams respectively.
- The resulting network graph is denoted by G = (V, E) containing the set of *communication, computation, source,* and *destination* nodes, and links respectively.
- Each communication link is characterized by a capacity (per flow unit), and a cost.
- While computational links are characterized by their computation and storage capacities (CPU/RAM).

5.1.2 Service Graph Modeling

A generic service can be described by a directed acyclic graph (DAG) R = (I, K) where vertices represent the service functions (e.g., stream processing operators), and edges correspond to data streams (or commodities).

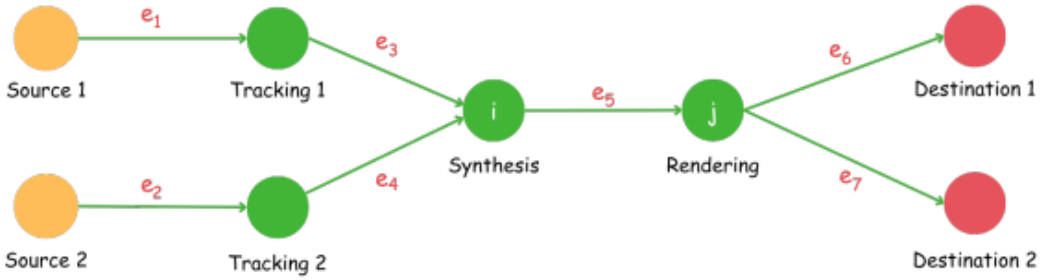

Figure 5. *Example of a service graph, where edges represent data streams (commodities) and vertices service functions. The example represents a NextG media application, in which streams from two sources go through tracking, synthesis, and rendering functions before being delivered to corresponding destinations.*

Fig. 5, shows an example of a general service graph (SG) where source commodities pass through a set of processing functions (Tracking, Synthesis, and Rendering) to finally multicast to their corresponding destinations.

The vertices with no incoming edges of the service graph represent source functions that produce source data streams (e.g., video capture), and the vertices with no outgoing edges represent destination functions that consume processed data streams (e.g., video display). Source/destination functions may also represent purely ingress/egress points injecting/ejecting data in/out of the network, and are always associated with a fixed and unique location in the cloud-network. At the same time, the remaining functions are subject to placement optimization.

5.2 *MuMeNet* simulator

Designing services for the MM is inherently more complex than a single-layer embedding problem. In classical VNE, one typically treats an incoming virtual-network request (VNR) as a static graph whose nodes and links must be mapped onto a substrate to optimize some objective (cost, revenue, utilization, etc.). By contrast, MM workloads demand a two-layer perspective:

- *Graph-layer:* At this level the concert (or more generally, the MM application) is modelled as a virtual service graph. The canonical question is: "How should we embed this graph onto the infrastructure to meet a given objective?" still applies, but with a crucial twist: the topology is dynamic. Audience members can join or leave at any moment, meaning the embedding must adapt continuously without violating stringent latency or QoS/QoE constraints.
- *Data Layer:* Above the graph mapping lies a data-centric layer that captures the real-time behavior of flows after they are placed. Classical VNE often idealizes the substrate, assuming the network can absorb jitter, packet loss, or congestion within a single link-capacity constraint. In MM scenarios, we cannot make that simplification. End-to-end musical interaction is hypersensitive to latency variation, packet loss, and even congestion spill-over from neighboring concerts or experiences. The data layer must, therefore, monitor and react to these runtime conditions e.g., by rerouting specific commodities, invoking redundant edge renderer, or dynamically adjusting bit-rates to maintain the illusion of co-presence.

To satisfy all these requirements, we need a framework that can consolidate all these variables in a single location. For this reason, we introduce ***MuMeNet***, a complete discrete-event simulation framework for musical metaverse scenarios. To meet QoS and QoE expectations, MM sessions must adapt in real time to changing network conditions and explicitly account for the concurrent multimodal traffic. These challenges underscore the need for novel modeling and simulation methods tailored to the unique aforementioned characteristics of MM sessions, along with specialized service provisioning

and orchestration strategies capable of capturing their real-time, interactive, and multicast-oriented nature.

To model, simulate, and analyze the problem of service provisioning in the MM, several aspects should be considered:

- *Scale and Heterogeneity:* A realistic MM session spans thousands of geographically dispersed users, multiple edge tiers, backbone transport, and cloud data centers. Recreating that hierarchy in hardware would demand a large number of servers, equipment, and investments.
- *Observability and repeatability:* Field measurements can reveal end-to-end delays, but make it difficult to expose what happens inside each compute node (e.g., buffers); nor can they be replayed under identical load. A simulation environment, by contrast, allows one to have deterministic control over every link, timestamp, and flow, enabling fair comparisons and statistically significant ablation studies.
- *Risk and iteration speed:* Early-stage policies often violate QoS constraints; running them live would produce audible glitches or broken sessions. The simulation environment offers a low-risk sandbox in which unsafe ideas can fail fast and guide the next design cycle.

5.2.1 *MuMeNet* vs existing Network Simulators

Packet-level network simulators: Their primary limitation is architectural; these tools assume static network topologies and treat packets as homogeneous byte streams, lacking the abstractions necessary to represent SGs or modality-specific latency constraints.

Cloud/Edge simulators: To explore service placement and resource allocation policies, the research community has widely adopted the CloudSim family of simulators, whose internal models follow the classic Virtual Network Embedding (VNE) abstraction.

Although these frameworks excel at graph-layer optimization, they cannot express information-centered replication, mixed-cast flows, or millisecond-level latency variance, capabilities that MM workloads require. In addition, critical network dynamics, such as packet jitter, burst loss, and congestion effects (the primary causes of musical desynchronization), are rendered invisible. As a result, a placement strategy that appears optimal in simulation may prove unusable once the dynamics of a real-world network are considered.

MuMeNet fills this gap by (i) embedding latency-annotated SGs onto a live packet topology, (ii) tracking audio, video, and haptic objects separately to compute inter-stream skew at micro-second resolution, and (iii) injecting bursty audience events that reshape traffic in real time. These capabilities enable faithful evaluation of orchestration algorithms under the exact timing constraints of musical co-presence.

5.2.2 *MuMeNet: Architecture and Design*

We now present the architecture of MuMeNet in more detail. Specifically, MuMeNet allows researchers to (i) develop flexible algorithm to embed MM dynamic SGs on arbitrary network topologies with real-world parameters, (ii) stream fine-grained traffic commodities through those embeddings under realistic link-level impairments, and (iii) measure, study, and analyze end-to-end musical-quality, QoE, and QoS for users with a large set of extensible evaluation metrics.

The simulator follows five practical design principles:

Modularity: Keeps each part: Network model, service-graph engine, traffic generator, and log collector, its module does changes in one place, without breaking the rest.

Extensibility: New algorithms and processing blocks can be added as plug-ins rather than rewriting core code. Through a central YAML/CLI interface, every key knob (e.g. capacity, rate) remains configurable, making large experiments straightforward.

Observability: The system provides comprehensive metrics and visualization (built-in dashboards and detailed event logs).

Simplicity: Only the essential features are implemented, ensuring the code is easy to read now while leaving clear hooks for future expansion.

Our simulator consists of two modules: *Core Layer* and *Scenarios Workspace*, the former is the principal part of the simulator responsible for the creation, management, and interactions between components. It groups all functionality that is common to every experiment. Scenarios Module, a directory module where users (developers) can create their customized scenarios and use cases. Each scenario in the scenarios module is treated independently.

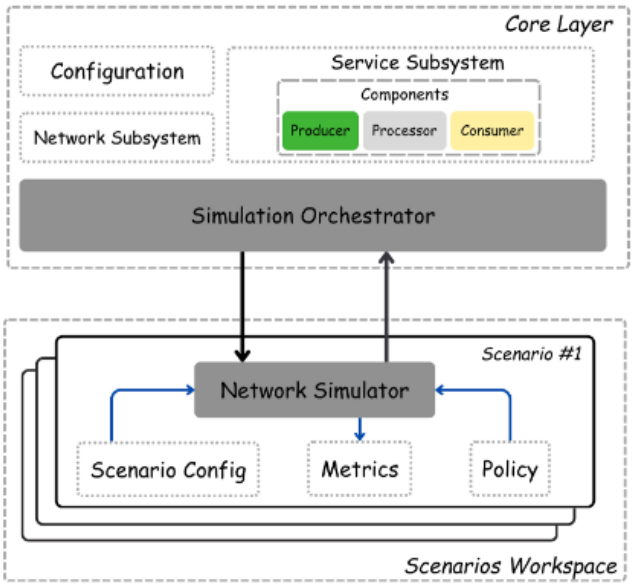

Figure 6. *MuMeNet architecture.*

The Core Layer is the invariant runtime backbone of MuMeNet. It collects every (simulation) service that must remain stable across experiments (e.g., configuration management, network modeling, component life-cycle control, and the discrete-event orchestrator) into a single, version-controlled module. It bundles the following four subsystems:

Simulator Configuration: This subsystem provides the authoritative source for all runtime parameters needed by the simulator. At start-up, the engine instantiates a BaseConfig object that pulls a version-controlled YAML/JSON file with default values (e.g., simulation horizon, time step, random seeds, logging levels, and directories), validates every field against a schema and reports any mismatch, then archives the resolved configuration alongside the run outputs to preserve full reproducibility.

This architectural approach facilitates extensibility through inheritance-based configuration customization. Researchers can derive domain-specific configuration classes from the BaseConfig base class within their respective scenario directories. Such derived classes retain the capability to override existing attributes or introduce additional parameters while preserving the underlying validation and persistence mechanisms. Upon initialization, the simulation engine maintains an immutable reference to the configuration object, which serves as the canonical parameter source for all core subsystems, including the network model, metrics collection framework, and policy plug-in architecture. This design pattern ensures centralized parameter management with strong type safety guarantees while simultaneously enabling scenario-specific customization without requiring modifications to the core system implementation.

Network Subsystem: This subsystem is the bridge between the abstract network graph template and the event traffic that flows through the simulator at run time. It performs two
high-level roles:

- Graph Construction: Assembling a topology whose nodes and links carry detailed resource attributes (e.g., packet loss, jitter, propagation delay). Moreover, a user can easily add some other dimensions.
- Topological Consistency Checking: Ensuring that the constructed topology obeys domain-specific constraints so that no unrealistic artifacts bias the results

Note that at the start of every run, the network subsystem ingests either a ready-made template derived from real deployments or a synthetic topology.

Service Subsystem: A SG is realized as an ordered set of components. Each component encapsulates the internal logic of a single service-graph node and is instantiated as one of three canonical roles. Custom behavior is added by subclassing the role-specific bases (Producer, Processor, or Consumer), while the shared interface keeps every extension drop-in compatible with the framework. A component in the simulator passes through a set of different phases. First, the developer instantiates the object, during this phase the simulator validates the provided metadata (e.g., I/O rates, state size, latency budget) against domain constraints. Second, the component is added to the global registry, making it discoverable by other elements. Finally, after having all required components successfully registered, the simulator connects them via provisional data-flows and exchanges lightweight "dummy" messages to verify type compatibility, buffer sizes, and back-pressure signals. Any mismatch triggers a descriptive error before the actual experiment begins.

Simulation Orchestrator: The Orchestrator constitutes the control plane of the simulator. Its design is deliberately centered on a discrete-event scheduling paradigm, because audience interaction workloads exhibit highly uneven traffic bursts and tight latency budgets that are best captured by event-driven, rather than fixed-timestep execution.

The orchestrator fulfills the following responsibilities:

- Global Time Monitoring: It moves the simulation clock to the timestamp of the next event in the queue.
- Event Dispatcher: It invokes the callback attached to each event, thereby triggering computation, communication, or rendering inside the appropriate component. The event also accepts a set of optional arguments if required.
- State Monitor and Update: It updates link capacities, queue lengths, and component states after every callback, guaranteeing that subsequent events observe a consistent world view.
- Variability Engine: It injects stochastic effects (propagation-delay distributions, packet-loss bursts, device jitter) at the precise moment they materialize. This information is initially provided by the user in the configuration files.

Every occurrence in the simulator such as, e.g., sending a packet, completing a CPU task, finishing an audio render, is encoded as an event object with three immutable fields: unique ID, timestamp (at which the event must start executing), and callback (the function to execute). The figure below provides a more in-depth view of the core engine, where each module/subsystem is shown along with its parts.

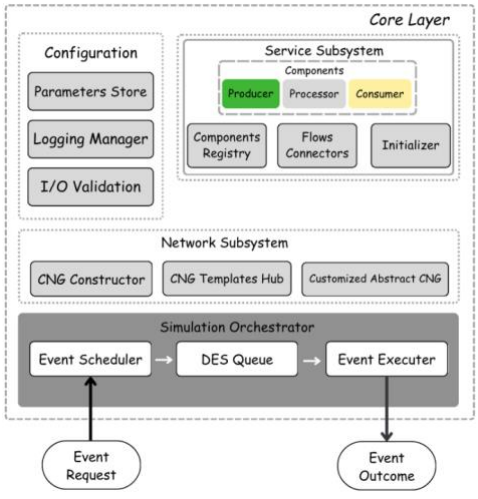

Figure 7. *The core layer in MuMeNet.*

This workspace allows users to create their customized scenarios Four modules compose the workspace, as detailed below.

Scenario Configuration Module: This module can extend the set of initial configuration parameters defined in the core module. The reason behind this design is to ensure separability: only the necessary parameters are called; otherwise, they remain set to default. In addition to this, we store the necessary scenario configuration in this module. We share with the user the set of possible parameters to consider for the pre-defined scenarios; however, if the user wants to create a completely customized scenario out of the provided template, this may require additional steps, as values have to be defined (and validated) first in the core part.

Policy Module: This module contains a set of abstract classes and methods to be extended for easy design of a customized policy. The policy's objective is to determine placement, routing, and resource allocation for the incoming SGs over the CNG, following static and dynamic approaches.

Metrics Module: This module collects real-time metrics of all components, monitors the changes of their statuses, and updates the list of metrics. This list depends on the use case; however, users can easily extend the base classes provided in this module to create any customized metric along with the interval of measuring or the set of components to use for collecting results.

Network Simulation Module: The scenario's key subsystem bridges the scenario with the core layer: it runs the orchestrator's network-level callbacks and returns updates or event requests, acting as the simulator's central communication hub, where all connections inside the simulator meet, from the CNG to the SG, and where metrics are collected before being sent to the "Metrics" module.

6. Future Developments and Research Directions

Based on the conducted analaysis of the state of the art in networking, we have identified a set of future developments, as detailed hereinafter.

AI-Driven Orchestration and Resource Allocation: Machine learning, particularly RL, can optimize resource allocation by dynamically adjusting network slices, prioritizing flows, and managing MEC offloading in real time. In MM settings, RL agents could anticipate congestion and preemptively adjust routing, reducing packet loss during critical performance segments.

Digital Twins for Performance Optimization: Digital Twins of performers, instruments, and venues can act as predictive models, simulating network load and rendering requirements before the actual performance. This enables pre-optimization of slices and MEC resources, reducing the risk of latency spikes during live events.

Beyond 5G (6G and Advanced Connectivity): 6G will likely bring sub-1 ms latency, THz communication, and AI-native architectures. These could enable fully holographic, multi-sensory MM experiences with massive participation. Early research in Reconfigurable Intelligent Surfaces suggests the ability to dynamically optimize radio propagation in real time for complex stage environments.

Standardization and Interoperability: The MM ecosystem would benefit from open APIs, unified QoS parameters, and standardized orchestration interfaces to ensure cross-platform compatibility. Current gaps in interoperability remain a barrier to large-scale adoption.

Inclusivity and Accessibility in MM Design: Future MM systems should integrate accessibility features, enabling participation for users with disabilities through adaptive XR interfaces, brain-computer interaction, and AI-assisted performance tools.

7. Conclusion

This deliverable presented the outcomes of the investigation into the potentialities and limitations of 5G networks in supporting the MM a novel ecosystem combining networked music performance, XR, and intelligent orchestration for real-time, immersive musical experiences. Building upon the three articles *"Towards a Networking Architecture for the Musical Metaverse: The MUSMET Vision"*, *"Architecting the Musical Metaverse: Lessons from 5G and Emerging Technologies"*, and "*MuMeNet: A Network Simulator for Musical Metaverse Communications"*, (both accepted in the 6th IEEE International Symposium on the Internet of Sounds), this document consolidates both the theoretical framework and experimental insights, providing a roadmap for future developments within the MUSMET project.

Our analysis confirms that 5G's core capabilities, including the support for the URLLC, eMBB, and mMTC use cases, Network Slicing, and MEC, are essential for achieving the stringent latency, reliability, and throughput demands of professional musical collaboration. Private SA deployments augmented with MEC demonstrated end-to-end latencies below 25–30 ms and reliability above 99.999%, supporting real-time multimodal exchanges across audio, video, and haptic channels.

Nevertheless, the study also reveals critical bottlenecks in current public 5G infrastructures, particularly in NSA networks that still depend on 4G cores, resulting in latency spikes, jitter, and uplink asymmetry. Incomplete slicing implementation, limited MEC coverage, and interoperability issues among vendors further constrain scalability and stability for large-scale MM deployments. To address these challenges, this deliverable introduces an AI-driven orchestration framework that leverages RL and DT technologies to dynamically allocate network, compute, and perceptual resources.

Finally, the deliverable outlines future directions including 6G, AI-native architectures, and Reconfigurable Intelligent Surfaces and introduces MuMeNet, a novel simulation framework for evaluating next-generation musical metaverse services. Together, these contributions provide a clear path toward a globally accessible, inclusive, and musically expressive metaverse ecosystem, aligning with MUSMET's vision of advancing Europe's leadership in immersive media and creative technologies.